



Ancient genomics

Der Sarkissian, Clio; Allentoft, Morten Erik; Avila Arcos, Maria del Carmen; Barnett, Ross; Campos, Paula F.; Cappellini, Enrico; Ermini, Luca; Fernandez Garcia, Rut; Rodrigues da Fonseca, Rute Andreia; Ginolhac, Aurélien; Hansen, Anders Johannes; Jónsson, Hákon; Korneliussen, Thorfinn Sand; Margaryan, Ashot; Martin, Michael David; Moreno Mayar, José Victor; Raghavan, Maanasa; Rasmussen, Morten; Sandoval Velasco, Marcela; Schroeder, Hannes; Schubert, Mikkel; Seguin-Orlando, Andaine; Wales, Nathan; Gilbert, M. Thomas P.; Willerslev, Eske; Orlando, Ludovic Antoine Alexandre

Published in:

Philosophical Transactions of the Royal Society B: Biological Sciences

DOI:

[10.1098/rstb.2013.0387](https://doi.org/10.1098/rstb.2013.0387)

Publication date:

2015

Document version

Publisher's PDF, also known as Version of record

Document license:

[CC BY](#)

Citation for published version (APA):

Der Sarkissian, C., Allentoft, M. E., Avila Arcos, M. D. C., Barnett, R., Campos, P. F., Cappellini, E., Ermini, L., Fernandez Garcia, R., Rodrigues da Fonseca, R. A., Ginolhac, A., Hansen, A. J., Jónsson, H., Korneliussen, T. S., Margaryan, A., Martin, M. D., Moreno Mayar, J. V., Raghavan, M., Rasmussen, M., Sandoval Velasco, M., ... Orlando, L. A. A. (2015). Ancient genomics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660), [20130387]. <https://doi.org/10.1098/rstb.2013.0387>

Review



Cite this article: Der Sarkissian C *et al.* 2015 Ancient genomics. *Phil. Trans. R. Soc. B* **370**: 20130387.
<http://dx.doi.org/10.1098/rstb.2013.0387>

One contribution of 19 to a discussion meeting issue 'Ancient DNA: the first three decades'.

Subject Areas:

evolution, genomics

Keywords:

ancient DNA, genomics, next generation sequencing

Author for correspondence:

Ludovic Orlando
e-mail: lorlando@snm.ku.dk

Ancient genomics

Clio Der Sarkissian, Morten E. Allentoft, María C. Ávila-Arcos, Ross Barnett, Paula F. Campos, Enrico Cappellini, Luca Ermini, Ruth Fernández, Rute da Fonseca, Aurélien Ginolhac, Anders J. Hansen, Hákon Jónsson, Thorfinn Korneliussen, Ashot Margaryan, Michael D. Martin, J. Víctor Moreno-Mayar, Maanasa Raghavan, Morten Rasmussen, Marcela Sandoval Velasco, Hannes Schroeder, Mikkel Schubert, Andaine Seguin-Orlando, Nathan Wales, M. Thomas P. Gilbert, Eske Willerslev and Ludovic Orlando

Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

The past decade has witnessed a revolution in ancient DNA (aDNA) research. Although the field's focus was previously limited to mitochondrial DNA and a few nuclear markers, whole genome sequences from the deep past can now be retrieved. This breakthrough is tightly connected to the massive sequence throughput of next generation sequencing platforms and the ability to target short and degraded DNA molecules. Many ancient specimens previously unsuitable for DNA analyses because of extensive degradation can now successfully be used as source materials. Additionally, the analytical power obtained by increasing the number of sequence reads to billions effectively means that contamination issues that have haunted aDNA research for decades, particularly in human studies, can now be efficiently and confidently quantified. At present, whole genomes have been sequenced from ancient anatomically modern humans, archaic hominins, ancient pathogens and megafaunal species. Those have revealed important functional and phenotypic information, as well as unexpected adaptation, migration and admixture patterns. As such, the field of aDNA has entered the new era of genomics and has provided valuable information when testing specific hypotheses related to the past.

1. The impossible genome

Ancient DNA (aDNA) research is full of surprises. Less than a decade ago, most experienced aDNA researchers believed that full genome sequencing of extinct species such as the woolly mammoth and Neandertals was impossible. The best available technology at the time was incredibly demanding in terms of fossil material, experimental work load and cost. First, each piece of target genomic DNA had to be amplified several times by PCR, then ideally PCR amplicons had to be propagated using bacterial vectors, and a number of clones had to be sequenced before a consensus sequence devoid of sequencing errors could be generated [1]. Furthermore, this whole procedure often needed to be replicated in another laboratory, before DNA sequences could be considered authentic [2].

The size of PCR amplifiable fragments was most often limited to approximately 100–150 base pairs (bp) at best, which represented little sequence information. With exceptionally well-preserved samples, the characterization of the whole approximately 16.5 kilobases (kb) of mitochondrial genomes [3–5] could be achieved using overlapping amplicons [6], but nuclear markers [7,8] were more difficult to amplify owing to their lower copy number per cell. Considering usual aDNA concentrations, each microlitre of DNA extract yielded one PCR amplicon at best. Therefore, it was generally necessary to destructively sample large amounts of fossil material to sequence complete mitochondrial genomes. The sequencing of the cave bear mitochondrial genome required for instance

1 g of bone material and not less than 570 PCR amplicons [5]. Assuming approximately 3 gigabases as the size of the nuclear genome and similar PCR success rates for mitochondrial and nuclear templates, this technology would have required approximately 180 kg of material and more than 103 million amplicons to generate a first draft of the cave bear genome. Given standard PCR and sequencing times, even on platforms with the highest throughput at the time (384 reactions per run), this would have required approximately 48 000 years of experimental work, excluding the time required for cloning! The characterization of even the best-preserved woolly mammoth specimens would have required similar efforts [4]. Two-round multiplex reactions [9], whereby a set of PCR targets are co-amplified from the same microlitre of DNA extract, could help reduce material and time requirements by one or two orders of magnitude [10,11], but operational costs would still amount to billions of US dollars. In summary, this technology limited palaeogenomics to the shorter ancient microbial genomes [12].

2. Shotgun sequencing of the first ancient mammalian genome

An alternative approach consisted of shotgun sequencing following aDNA ligation into bacterial plasmids [13,14]. This could be done at large sequencing centres, though with two major limitations. First, most of the sequences generated in fact do not originate from the organism of interest, but from environmental microbes that colonize the tissue after deposition. Therefore, no more than 26.9 kb of the cave bear genome could be reconstructed with this approach from a total of 14 027 sequences [13]. The second limitation was the heavy experimental load required for bacterial cloning. The invention of ‘emulsion PCR’, whereby each DNA library template is amplified in a water–oil emulsion droplet, and the development of the 454 platform provided a time-effective alternative to bacterial cloning by processing hundreds of thousands of sequencing reactions in parallel [15]. Applied to DNA extracts from an approximately 28 000 years (28 kyr) old mammoth bone sample, this technology provided, in a 6 h long run, 28 megabases (Mb) of metagenomic data, of which approximately 13 Mb belonged to the mammoth genome [16]. This demonstrated for the first time that sequencing of complete mammalian genomes was probably achievable from realistic amounts of bone material.

The field improved further, with the realization that hair constitutes a remarkable source of high-quality aDNA [17,18] that could be subjected to efficient decontamination procedures [19] inapplicable to bones. Deep sequencing of DNA from ancient mammoth hair yielded approximately 80% of sequences identified as being of mammoth origin [20], thus providing a first draft covering approximately 70% of the mammoth genome, with an overall sequencing error rate estimated at 0.345%. These data revealed that 99.4% of the sequenced mammoth genome was identical to the African elephant genome and identified 29 mammoth genes with specific non-synonymous mutations of potential functional importance. Testing this list of gene candidates following the methodology that revealed the association between an allelic variant at the MC1R gene and blond coat-colour [9] could illuminate our understanding of the genetic make-up of mammoths.

3. The first ancient human genome

By the time the first draft of the mammoth genome was characterized, new sequencing technologies with higher throughput were available [21]. The Illumina Genome Analyzer II platforms could generate 180 million sequence reads per run. This massive sequencing throughput, combined with the high endogenous DNA content of hair, and preservation in a cold environment made the sequencing of the first ancient human genome possible. The individual sequenced was a palaeo-Eskimo belonging to the Saqqaq culture, who lived along the southwestern coast of Greenland 4 kyr ago [22]. The sequence information gathered represented an average depth of 20-fold across 79% of the genome and led to the identification of a large catalogue of high-confidence single nucleotide polymorphisms (SNPs), some of which not only confirmed his hair colour but also showed that he was of the A+ blood type and most likely had brown eyes, dry earwax, as well as a metabolism and body mass index adapted to cold climate. ADMIXTURE [23] and principal component analysis [24] of the SNP information indicated no affinity with modern-day Europeans, thus ruling out possible contamination, a problem that had plagued ancient human DNA studies for decades [25], as well as attempts at sequencing the Neandertal genome a few years earlier [26,27]. Analyses also revealed a much closer genetic affinity with contemporary Chukchis and Koryak populations of northeast Siberia, than with present-day Greenlandic Inuit. Divergence times with the Chukchi population closely matched the radio-carbon date for the ancient individual, suggesting that the Saqqaq ancestors entered Greenland soon after they separated from their Old World relatives and were later replaced by the ancestors of modern-day Inuits. This study demonstrated the immense potential of palaeogenomics towards reconstructing the population history of humans in much greater detail than what can be achieved from patterns of modern genomic variation alone.

A recent re-analysis of the Saqqaq sequences also revealed epigenomic signatures indicative of gene expression. Sequence depth variation showed a strong approximately 200 bp periodicity, which is characteristic of the length of one nucleosome and spacer block [28]. Within a shorter range, a 10 bp periodicity corresponding to the size of a DNA helix turn was also detected, reflecting preferential cleavage of the DNA backbone that faces away from nucleosome protection. Importantly, patterns of sequence depth variation at known nucleosome arrays showed strong correlations with nucleosome occupancy. It therefore seemed that in addition to the DNA sequence, the compaction state of the chromatin could survive in fossils, and that variation in read depth, corrected for base compositional bias, could be used as a footprint of nucleosome protection to reconstruct genome-wide nucleosome maps. Regional methylation levels could also be tracked in the Saqqaq genome sequence owing to the fact that it had been obtained by amplifying DNA libraries using a polymerase that amplifies cytosines deaminated due to post-mortem damage only when methylated [29–31]. Focusing on read starts, where deamination rates are highest, estimated methylation levels were found to recapitulate known genomic patterns at different classes of CpG promoters, splice sites and CTCF transcriptional repressor sites. Strikingly, the Saqqaq methylome also appeared closer to that of modern hair than to other somatic tissues. As nucleosome occupancy and cytosine methylation influence gene

expression, those could be used to predict gene expression levels in the Saqqaq hair cells. As expected, key structural components of hair, such as keratins and trichohyalin, were predicted as highly expressed, demonstrating that ancient gene expression levels can be gathered directly from ancient sequence data, even in the absence of RNA. This approach can therefore complement functional SNP genotyping [32] and proteomics [33] to gather functional information from ancient individuals and investigate the dynamics and evolutionary significance of epigenomic changes.

4. Ancient anatomically modern humans

Recent mixed ancestries among modern human groups can limit our ability to infer their true past population history. Following the sequencing of the Saqqaq genome [22], a variety of ancient human genomic studies have shed light on major events in the human population history [34–41]. Of key importance, post-mortem DNA damage patterns and heterozygosity levels observed in the mitochondrial genome [42], or on the X-chromosome of male individuals, have provided robust approaches to rule out modern contamination.

Using 600 mg of hair collected in the 1920s from an Aboriginal Australian male, Rasmussen *et al.* [34] were able to reconstruct his genome at 6.4-fold coverage and found no evidence for recent European admixture or contamination. Genomic affinity was revealed with present-day Aboriginal Australians, as well as Bougainville and New Guinea Highland Papuans. More importantly, being the first Aboriginal Australian genome sequenced, this dataset allowed for the testing of two alternate models of modern human dispersal into eastern Asia. A statistics based on quartet genome alignments (D_{4P}) showed an excess of genomic sites in support of a population tree where Aboriginal Australians and Africans cluster together, separate from Europeans and Asians. Population split times suggested that Aboriginal Australians separated from the ancestral Eurasian population around 62–75 kyr ago, which in turn radiated into the European and Asian branches some 25–38 kyr ago. These results favour the hypothesis of the ‘multiple dispersal’ model [43], which was also supported by the excess of shared derived alleles observed between Asians and Aboriginal Australians, reflecting population migration from the mainland. By contrast, the Australian Aboriginal genome lent little support to the alternative ‘single dispersal’ model proposing that humans expanded out of Africa into Eurasia 50 kyr ago [44] through a series of founder events, which ultimately gave rise to the colonization of Australia and the diversification of Aboriginal Australian populations.

Ancient human genomic data has also shed light on another fiercely debated topic in anthropology, namely the peopling of the Americas. Genomic signatures of an Upper Palaeolithic (approx. 24 kyr ago) male juvenile excavated at the Mal’ta site, south central Siberia, Russia revealed no strong connection with present-day eastern Asians [39]. Tree-based analyses of population splits and admixture events (TREEMIX [45]) instead placed the Mal’ta specimen basal to western Eurasians and identified gene flow from Mal’ta to Native Americans. Shotgun sequencing of another individual from the same region and dating to post-last glacial maximum (LGM; 17 kyr ago) showed a strong affinity with the Mal’ta specimen, suggesting that the population’s gene pool was rather stable during the LGM, and consequently, that the

populations from the region changed within the last 17 kyr. The western Eurasian component of the Mal’ta specimen suggests that Upper Palaeolithic populations ancestral to present-day western Eurasians had a distribution range that extended further northeast. This is consistent with the discovery of a number of anthropomorphic Venus figurines at the Mal’ta site, reminiscent of Upper Palaeolithic sites in western Eurasia. Interestingly, no particular genetic affinity was detected between present-day western Eurasians and a 40 kyr old individual excavated at the Tianyuan cave in north east China [38], confirming that the results were not affected by recent events in the population history of modern eastern Asians. The contribution of the Mal’ta lineage to the Native American gene pool gives further support to the hypothesis of a Siberian origin for present-day Native Americans, and the gene flow between Mal’ta and the ancestors of 52 Native American populations from Greenland to southern Chile was estimated to be responsible for 14–38% of the current Native American ancestry. This gene flow occurred before 12.6 kyr, which is the age of a child excavated at the Anzick site, Montana, USA, whose genome also revealed a ‘Mal’ta-like’ component [40]. The Anzick child belonged to the Clovis culture, the oldest archaeological complex in North America, and was part of a meta-population directly ancestral to all contemporary Native Americans outside of Canada and the Arctic. Overall, it appears that the ancestors of present-day Native Americans were the descendants of at least two population backgrounds, one related to the Mal’ta individual, showing a western Eurasian affinity, and another related to present-day eastern Asians, as suggested by the strong eastern Asian genetic component found among Native Americans [46]. This new model for the origin of Native Americans potentially solves the mystery surrounding the presence of non-east Asian morphological features in the skulls of the first Americans.

Ancient genomics has also provided invaluable clues to understand the complex genetic make-up of Europeans. The genome of a 5.3 kyr old Copper Age ‘Tyrolean Iceman’ revealed genetic discontinuity with current inhabitants of the Alps, with the Iceman showing a greater genetic affinity with southern European populations, and in particular with Sardinians [35]. This finding suggests that the current Sardinian population represents a remnant of an ancient and previously more widespread component of the European gene pool. Similarly, genome-wide data indicated that an approximately 5 kyr old early farmer from Sweden was more closely related to southern Europeans than to present-day northern Europeans and three contemporary hunter-gatherers of Sweden [36]. Those three hunter-gatherers as well as two 7 kyr old hunter-gatherers from the Spanish cave called La Braña [37,41] fell outside the present-day European genomic diversity but exhibited a closer genetic affinity with present-day populations of northern Europeans. One La Braña individual [41] was found to share ancestry with the Siberian Mal’ta individual, thus providing further evidence for the genetic and cultural links between the West Eurasian Mesolithic and the Siberian Upper Palaeolithic. All together, these results suggest a shared genomic background of hunter-gatherers across North Eurasia, as well as a migration-driven transition associated with the advent of the agricultural lifestyle in Europe.

Genomic data from ancient humans in Europe revealed information about their likely phenotypes and health status.

For instance, the Tyrolean Iceman probably had brown eyes, belonged to the O+ blood group and was lactose intolerant. He was also homozygous for alleles associated with major risks for coronary heart disease and atherosclerosis, and infected with *Borrelia burgorferi*, the pathogen responsible for Lyme disease [35]. La Braña hunter–gatherers probably had difficulties digesting milk and starch, were dark haired and dark skinned and had non-brown eyes. The derived variants of immunity genes found in the La Braña genome suggested that hunter–gatherers were adapted to resist multiple types of infection that were commonly believed to have emerged much later with the advent of the agricultural lifestyle [41].

5. The genomics of archaic hominins

The Neandertal genome project represents a milestone in ancient genomics, as it led to major technical improvements, both for generating and analysing aDNA data [47]. Most of the final sequences used for the first genome draft assembly required no more than 400 mg of bone material sampled from three female specimens excavated at the Vindija cave, Croatia and dated to 38–44 kyr ago [47]. More recently, Prüfer *et al.* [48] generated a high-quality genome from a female Altai Neandertal from the Denisova cave, Russia, and a low-coverage draft genome from an approximately 60–70 kyr old Neandertal infant. The high-quality genome could be obtained owing to a combination of exceptionally high fraction of reads aligning to the human genome and minimal contamination levels. This vast genomic dataset makes the Neandertals the best-characterized extinct species today. The high-quality Altai genome revealed high inbreeding coefficients compatible with half-sibling mating, and temporal variations in the Neandertal population size, which was estimated to have been about a tenth of that of present-day humans, despite a broad Eurasian geographical range extending from the Iberian Peninsula to the Altai mountains.

Earlier genetic screening of a finger bone excavated at the Denisova cave had revealed the presence of an archaic hominin belonging to a mitochondrial lineage very distinct from modern humans and Neandertals [49]. Enzymatic treatment prior to DNA library preparation eliminated the vast majority of nucleotide misincorporations resulting from post-mortem damage, and further, employing paired-end sequencing and collapsing mate reads that showed sufficient overlap delivered a first draft of the nuclear genome with limited error rates [50]. The nuclear sequence data supported a different population scenario than the mitochondrial data. The archaic hominin appeared indeed to belong to a group distinct from both modern humans and Neandertals, but more closely related to Neandertals than to modern humans. This group was named Denisovans after the cave where it was first discovered. A molar tooth excavated at the Denisova cave contained enough endogenous DNA to reconstruct another full mitochondrial sequence using target enrichment. The latter appeared closely related to the sequence from the finger bone. The development of a new DNA library preparation method targeting single-stranded molecules enabled the reconstruction of an ancient genome showing a quality comparable to that of modern genomes sequenced at similar depth [51].

The temporal limits of archaic human genomics were recently pushed back with the sequencing of the complete mitochondrial genome of a 400 kyr old hominin from the Sima de los Huesos cave in northern Spain [52]. The sequenced individual is morphologically characterized as *Homo heidelbergensis*, a lineage commonly considered as pre-Neandertal. Yet, the mitochondrial genome appeared closer to Denisovans than Neandertals. Further genetic information, at the nuclear level, is required before this mitochondrial affinity can be confirmed or this result can alternatively be demonstrated as a consequence of a complex population history involving incomplete lineage sorting and/or gene flow [53].

The Neandertal and Denisovan genomes have revealed important information regarding admixture among archaic hominins and anatomically modern humans. *D*-statistics [54] indicate that Neandertals shared an excess of derived alleles with non-African modern populations [47–48,50–51]. This suggests that anatomically modern humans and Neandertals admixed in Eurasia. According to the latest estimates [48], this gene flow introduced 1.5–2.1% of Neandertal ancestry (most closely related to the individual from the Caucasus than the Altai) into the genome of non-African individuals. However, it is still debated whether the genomic patterns observed result from admixture between anatomically modern humans and Neandertals [55,56] or reflect ancestral population structure in Africa [57,58]. In contrast to Neandertals, Denisovans showed no evidence of gene flow into most present-day Eurasian populations, but did contribute to the gene pool of modern Melanesians [51,59] and, to a lower extent, of mainland Asian populations [48,60]. The admixture signature in present-day Papuans is greater on the autosomes than on the X-chromosome, possibly indicating the presence of hybrid incompatibility alleles on the X-chromosome, or that the gene flow preferentially involved Denisovan males and human females. Archaic hominin populations also appear to have mixed with each other. The level of Neandertal gene-flow into Denisovans is currently estimated at more than 0.5% [48] and an additional gene-flow into Denisovans originating from an unidentified hominin population (representing an outgroup to modern humans, Denisovans and Neandertals) has been proposed.

The archaic hominin genomes have importantly helped narrow down the genetic changes that make us humans [47–48,50–51]. Our understanding of how those relate to phenotype is, however, still in its infancy. In Denisovans, one difference was found in *EVC2*, a gene whose mutated alleles cause wider dental pulp cavities and fusion of tooth roots, both of which are common in the teeth of archaic hominins [51]. In Neandertals, some genetic variants in the *RUNX2* gene have been linked to cleidocranial dysplasia, which is associated with bell-shaped rib cages and changes in dental morphology, all of which represent major phenotypic differences between Neandertal and modern humans [47]. Functional assays also showed that the microRNA mir-1304 might be one factor involved in the difference in tooth morphology between modern humans and Neandertals [61]. The availability of the genome sequence allows anyone interested in a particular locus to investigate the variants present in archaic hominins, and potentially discover advantageous alleles that some modern human populations acquired from archaic hominins. Many such examples are now described and concern genes that are almost exclusively

involved in the innate immune system (*STAT2* [62]; *OAS1* [63]; *HLA* [64]).

6. Towards Middle Pleistocene genomes and proteomes

With high-quality genomes from the Holocene and the Late Pleistocene in hand, the question soon became how far back in time palaeogenomics could be pushed. In 2012, deep sequencing of DNA extracts from a 110–130 kyr old bone delivered genome-wide information from a polar bear [65]. This suggested that, at least in cold environmental conditions, such as those found in the Arctic Ocean Svalbard archipelago where the polar bear material was discovered, palaeogenomics could break the Middle Pleistocene time barrier (125–781 kyr ago). At that time, the genetic evidence that DNA could survive over several hundreds of thousand years was rather scarce and limited to the pyrosequencing of no more than 16 bp of mitochondrial bases from approximately 400 kyr old cave bear specimens [66], PCR amplicon sequencing of minibarcodes from 450–700 kyr old ice cores [67] and 400–600 kyr old sediment cores [68]. Yet, successfully sequencing Middle Pleistocene genomes would provide much needed perspectives across a broad range of evolutionary biology questions. Not less than five archaic hominins were living during the Middle Pleistocene [69], including the most recent common ancestor of anatomically modern humans, Neandertals and Denisovans. The Middle Pleistocene also experienced numerous radiations and extinctions of fascinating megafauna lineages [70,71], as well as major climatic changes, involving the succession of many glacial and interglacial episodes, in contrast to only one for the Late Pleistocene. Clearly, pushing the limits of palaeogenomics to the Middle Pleistocene would represent a major step forward.

The empirical demonstration that such a step is possible came from a fragment of horse metapodial bone excavated in 2003 at Thistle Creek (TC), Yukon, Canada [72]. The specimen was found within a stratigraphic layer associated with the Gold Run tephra and dated to 735 ± 88 kyr BP, in agreement with palaeobotanical and micromammal fossil analyses, which also indicated an Early–Middle Pleistocene age [73–75]. The line of evidence suggesting that biomolecules, including DNA, could survive for such a long time included: (i) the detection of amino acids within the bone matrix by time of flight secondary ion mass spectrometry, (ii) the identification of the three most abundant amino acids in the primary sequence of collagen (glycine, proline and alanine) in the bone matrix, (iii) the direct sequencing of a variety of peptides representing 72 proteins from the bone matrix and the circulating blood, (iv) the presence of significantly greater levels of protein degradation by glutamine deamidation in the TC horse than in a younger Late Pleistocene Siberian mammoth, and (v) the estimation of considerably higher levels of DNA damage in the TC horse than in younger Late Pleistocene horses also preserved in the Arctic permafrost. Additionally, phylogenetic inference based on complete mitochondrial genomes revealed that the TC horse fell outside the range of genetic variation of modern and Late Pleistocene horses. This was confirmed using the full set of protein-coding nuclear genes and a total of eight other genomes sequenced for comparison, including a 43 kyr old horse [72]. In addition, the retrieval of genomic

information from Middle Pleistocene specimens is compatible with the long-term survival of DNA predicted by the empirical model of DNA degradation through time proposed in [76].

The TC horse genome was sequenced at approximately 1.1-fold coverage using a combination of second-generation (Illumina) and third-generation (Helicos) sequencing. The latter, based on true single DNA molecule sequencing (tSMS), appeared to be advantageous when targeting short and damaged molecules for several reasons. First, this technology is PCR-free and, thus, devoid of PCR-related bias. Second, it does not require extensive enzymatic manipulation or repeated DNA purification steps, thereby maximizing DNA recovery and reducing the risks of enzyme incompatibility with chemically modified ancient templates. Third, this technology operates with single DNA strands and from any available 3'-hydroxyl group available. Consequently, with higher densities of single-strand breaks compared with modern DNA, aDNA templates present a higher chance of being sequenced on this platform. Methodological improvements of the sequencing protocol [77,78] and the development of dedicated bioinformatics strategies to improve the sensitivity and accuracy of read alignment against the horse reference genome [79] were necessary to optimize the analysis of aDNA molecules using tSMS. Altogether, this resulted in the identification of 4.21% of Helicos reads as endogenous horse DNA versus only 0.47% for Illumina reads.

The TC horse genome sequence was used first to date the time of the most recent common ancestor of horses, donkeys, zebras and asses at approximately 4.0–4.5 million years (Myr), which corresponds to twice the age of the first widely accepted *Equus* fossil from the palaeontological record. This new calibration point provided a genome-wide mutation rate that was used for scaling the palaeodemographic profile reconstructed from high-quality modern diploid genomes following pairwise sequentially Markovian coalescent inference. The profile revealed three major periods of demographic expansions and contractions for horses within the last 2 Myr, the last of which was consistent with ecological niche modelling and palaeoenvironmental data showing grassland expansion prior to the LGM followed by a massive post-LGM range contraction [80]. Interestingly, Bayesian skyline reconstructions based on the ancient mitochondrial genomes sequenced in this study, as well as tip-calibration, showed similar demographic changes, providing an independent validation of the novel calibration point proposed for *Equus*.

More fundamentally, the sequence data provided a unique snapshot of aDNA molecules from the Middle Pleistocene, revealing for the first time the presence of 3' overhangs [77] and fragmentation levels compatible with the survival of ultra-short fragments (25 mers) over 1 Myr. As the latter provide sufficient information for mapping, environmental conditions close to those in place at TC should therefore enable the characterization of 1 Myr old genome [81]. Outside the Arctic, temperate caves that represent an environment with virtually no variation in temperature are also likely to offer preservation conditions compatible with the reconstruction of Middle Pleistocene genomes, as shown recently by the analysis of bone material from Sima de los Huesos at Atapuerca [52,82]. In these studies, the experimental procedure, which combined a newly developed extraction method tailored to the retrieval of ultra-short DNA fragments, single-strand Illumina DNA libraries [51,83] and target-enrichment capture, recovered enough high-quality DNA reads to reconstruct a near complete

mitochondrial genome of a Middle Pleistocene cave bear and *H. heidelbergensis* [52].

7. Non-mammalian palaeogenomics

Despite major advances in genomic technologies, assembling complete plant genomes is a major challenge even for modern samples owing to their large, highly repetitive and heterozygous genomes, confounded by varying ploidy-levels, even within genera. Thus small-scale plant aDNA studies have been undertaken on maize [84], barley [85], cotton [86], wheat [87] and bottle-gourd [88], revealing patterns of crop adaptation and migration (reviewed in [89]). Larger and more in-depth studies using ancient plant genomes are expected in the coming years given the economic importance of major crops and the possibility of reintroducing alleles involved at various stages of the domestication process. Herbarium collections hold great potential as resources for future investigations of historical genomics as the specimens are generally well preserved and often meticulously annotated. High endogenous DNA content has allowed the genomic characterization of 'ancient' genomes of the plant pathogen *Phytophthora infestans*, the oomycete responsible for the Irish potato famine [90–92]. Herbarium collections also hold potential for population genomics. Future studies on ancient plant pathogens should reveal more details about their coevolution with food crops and the history of their human-mediated migrations, potentially leading to insights for crop breeding and management. Furthermore, the study of aDNA from non-domesticates such as forest trees will probably shed light on changes in biodiversity during past climatic events [93].

It has also recently been shown that RNA is preserved in some ancient seeds, even better than aDNA in maize kernels [94], presenting an opportunity to directly test evolutionary changes in gene expression at a key developmental stage [94]. Comparative genomic approaches and selection scans will also probably narrow a series of candidate loci that were adaptive in a range of environmental conditions. It should also be stressed that new aDNA reservoirs, such as egg shells [95] and dental calculus [96,97], are constantly being discovered, and should greatly benefit from deep-sequencing approaches.

8. Ancient genomes: a user's manual

(a) DNA damage patterns as authenticity indicators

With the introduction of genomics, new authentication criteria for DNA have emerged, with one of the most essential being the detection of typical signatures of post-mortem damage. Using blunt-end ligation of double-stranded templates, cytosine deamination at 5'-overhangs results in greater C → T misincorporation rates towards sequence starts [98]. At read ends, this signature is converted into a complementary increase in G → A misincorporation rates. This typical post-mortem damage signature is modified depending on the molecular tools used during library building and amplification. AT-overhang ligation, for instance, was shown to introduce significant biases in the sequence composition of library inserts, which discriminates against templates starting with thymine residues [99] and, consequently, deaminated cytosines (molecular analogues of thymines). This not only reduces the molecular complexity

of DNA libraries, but also transforms the expected nucleotide misincorporation pattern, which peaks at the second position from sequencing termini. Likewise, DNA polymerases of the *Pfu* family, such as *Phusion*, cannot bypass uracil residues [22]. As a result, no increase in C → T misincorporation rates are detected at sequence starts [31], except at methylated CpGs [29].

Procedures based on single-stranded templates also show a different nucleotide misincorporation profile. With Helicos tSMS, a sequence reverse complementary to the original template strand is generated by extension from the blocking site (figure 1a). Cytosine deamination at 3'-overhangs of the template strand will, therefore, increase G → A instead of C → T misincorporation rates [77,78]. With single-strand Illumina DNA libraries [51,83], whenever inserts are paired-end sequenced or sequenced over their full length, the expected misincorporation pattern corresponds to an increase of C → T rates at both sequence starts and sequence ends (figure 1b).

One type of DNA damage pattern, where the genomic position located upstream of sequence starts is enriched in purines [98], appears to be common to all library building protocols. This probably reflects a mostly depurination-driven post-mortem DNA fragmentation process. Preferential loss of adenine over guanine residues has been observed for aDNA extracts younger than a century [100], but guanine residues are preferentially lost for much older material, suggesting two temporally independent depurination dynamics at adenine and guanine residues. The resonance structure present within guanine residues and reducing the activation energy required to break the bond with the deoxyribose might influence these dynamics [101].

Restricting analyses to the population of sequences exhibiting typical damage patterns has been shown to enable genuine data recovery, even in the presence of significant levels of contamination [39,52,102]. Nucleotide misincorporation and DNA fragmentation patterns can be detected using ad hoc programs such as the MAPDAMAGE software [103,104], and post-mortem degradation parameters can be quantified from read alignments against reference genomes [103,104]. In light of the versatility of the signatures described above, we recommend that the same molecular methods should be used when comparing DNA damage parameters across a range of samples.

(b) Limiting the impact of post-mortem DNA damage

Fitting a DNA damage model to the data can also be used to limit the impact of C → T and G → A misincorporations in downstream analyses. In particular, the confidence placed on any nucleotidic base along a sequence can be downscaled post-mapping according to the probability of the base being affected by post-mortem damage [104]. This approach was shown to reduce the false positive rate of SNP calls on the Saqqaq data [104]. Ideally, the damage model should be applied during the mapping step itself, in order to improve read alignment accuracy and sensitivity, as currently implemented in the programs MIA [105], ANFO [47] and sesam [22], but usage of these softwares has been limited mostly owing to long-running times (MIA) and the inability to handle indels (sesam).

For now, the most common strategy for limiting the impact of misincorporations in downstream analyses has consisted in a first authentication of the data based on a

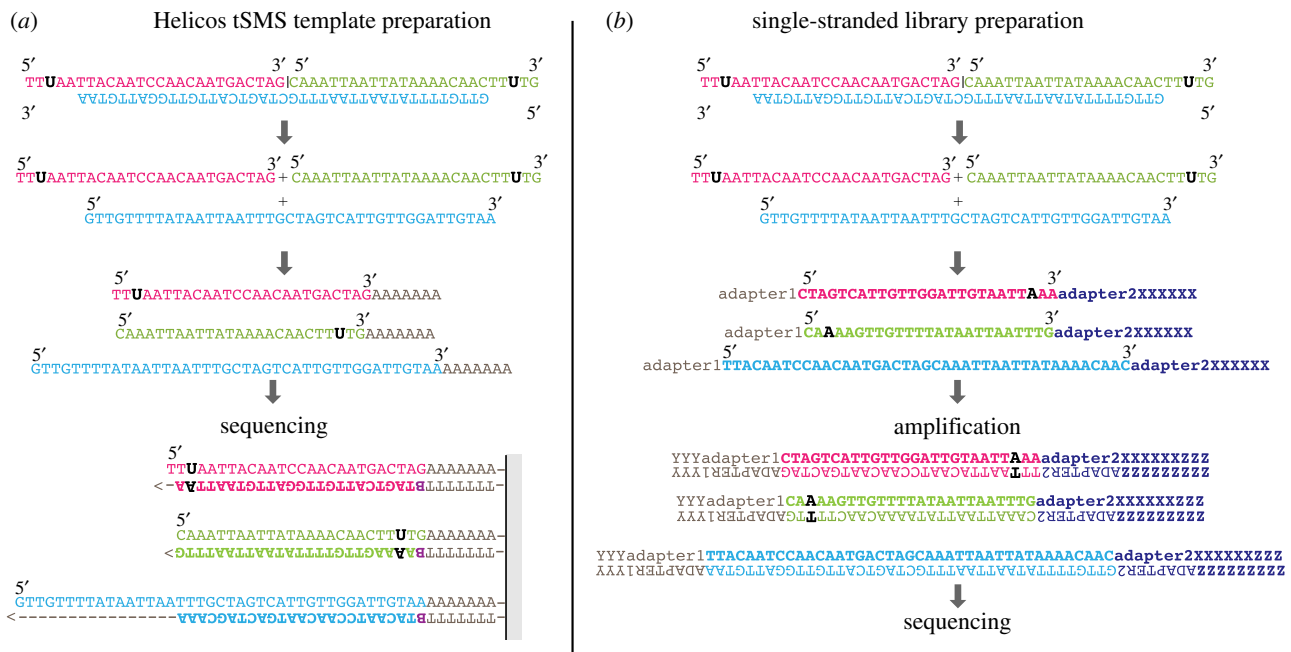


Figure 1. Accessing single-strand information from aDNA templates. Comparison of template preparation for (a) Helicos tSMS [77] and (b) single-stranded library building [51], in the case of a double-stranded DNA molecule showing a single-strand break ‘|’ and two cytosines deaminated to uracils (U). Both methods lead to three sequence reads, each represented by a different colour (pink, blue and green), thus providing strand information. Conversely, double-stranded molecules with a single-strand break built into double-stranded libraries lead to the sequencing of one read. Cytosine deamination in overhangs leads to an artefactual increase in G → A transitions in Helicos tSMS sequence reads and C → T transitions in sequence reads obtained from single-stranded libraries. (a) ‘B’, dCTP, dGTP or dATP virtual terminator. (b) ‘X’, ‘Y’ and ‘Z’ adapter/primer sequence.

small subset of sequences followed by a second production phase using nucleotide misincorporation-free libraries [47,106–108]. This is done by treating the aDNA extract with a cocktail of two enzymes, where the uracil-DNA glycosylase targets deaminated cytosines and generates abasic sites, which represent the targets for the EndoVIII endonuclease [29]. As a result, shorter DNA templates cleaved at damaged sites are ligated to adapters and incorporated into libraries. This strategy, in addition to high depth-of-coverage and paired-end sequencing where almost every single base position of the insert is read twice, has been essential for generating the high-quality Denisovan genome sequence [51].

(c) Increasing the relative amount of target DNA I: ancient DNA extraction

While standard silica-based extraction procedures [108] are biased towards molecules longer than approximately 40 bp, a recent extraction method has been developed to target the ultra-short fraction of DNA extracts [82]. Assuming that DNA fragmentation follows a one-order kinetics, the amount of aDNA available decreases exponentially with fragment size [76,109]. By targeting short DNA fragments, the new extraction method should thus drastically increase the amount of aDNA material in extracts, thereby increasing the molecular complexity of DNA libraries. This method is an important advance towards the sequencing of significantly older DNA templates, as well as DNA from environments offering poor preservation conditions. Methods have been devised for the preferential extraction of DNA from molecular preservation niches that can be found within fossils. Such niches have been proposed to correspond to crystal aggregates present in the most interior parts of bones where endogenous DNA is protected from hydrolysis and microbial invasion [110]. One

promising method has shown great success with permafrost preserved bone material, such as mammoth [111] and horses [77,78,112]. This method involves a first partial digestion of the bone powder, before undigested bone pellets are recovered and digested a second time in a fresh buffer. Pairwise tests have shown higher endogenous contents in the extract prepared from the second digest, as well as lower levels of cytosine deamination [77,78,112] and fragmentation [78]. If confirmed by additional tests on a range of ancient samples, targeting such molecular preservation niches could significantly reduce the costs related to ancient genome sequencing.

(d) Increasing the relative amount of target DNA II: ancient DNA library construction and amplification

Apart from library construction methods [99], the type of DNA polymerase used also significantly impacts the complexity of amplified DNA libraries [113]. Standard polymerases for aDNA research, such as *Taq* Gold, significantly skew the size distribution and base composition of the pool of molecules amplified towards short and GC-rich templates, which limits the ability to sequence the entire molecular diversity originally present in the DNA library. DNA polymerases can also be blocked during library amplification owing to the presence of atypical bases in aDNA templates [114], which can in turn provide an advantage to non-modified modern contaminants. The exact amount of such blocking DNA lesions is still largely unknown but would represent 10–40% of library templates according to a recent estimate using a limited number of fossil specimens from permafrost and temperate caves [114]. Devising methods for repairing such templates and/or capturing preferentially damaged templates [53] could further improve accessibility to aDNA molecules. DNA polymerases capable of bypassing damage lesions in DNA molecules have

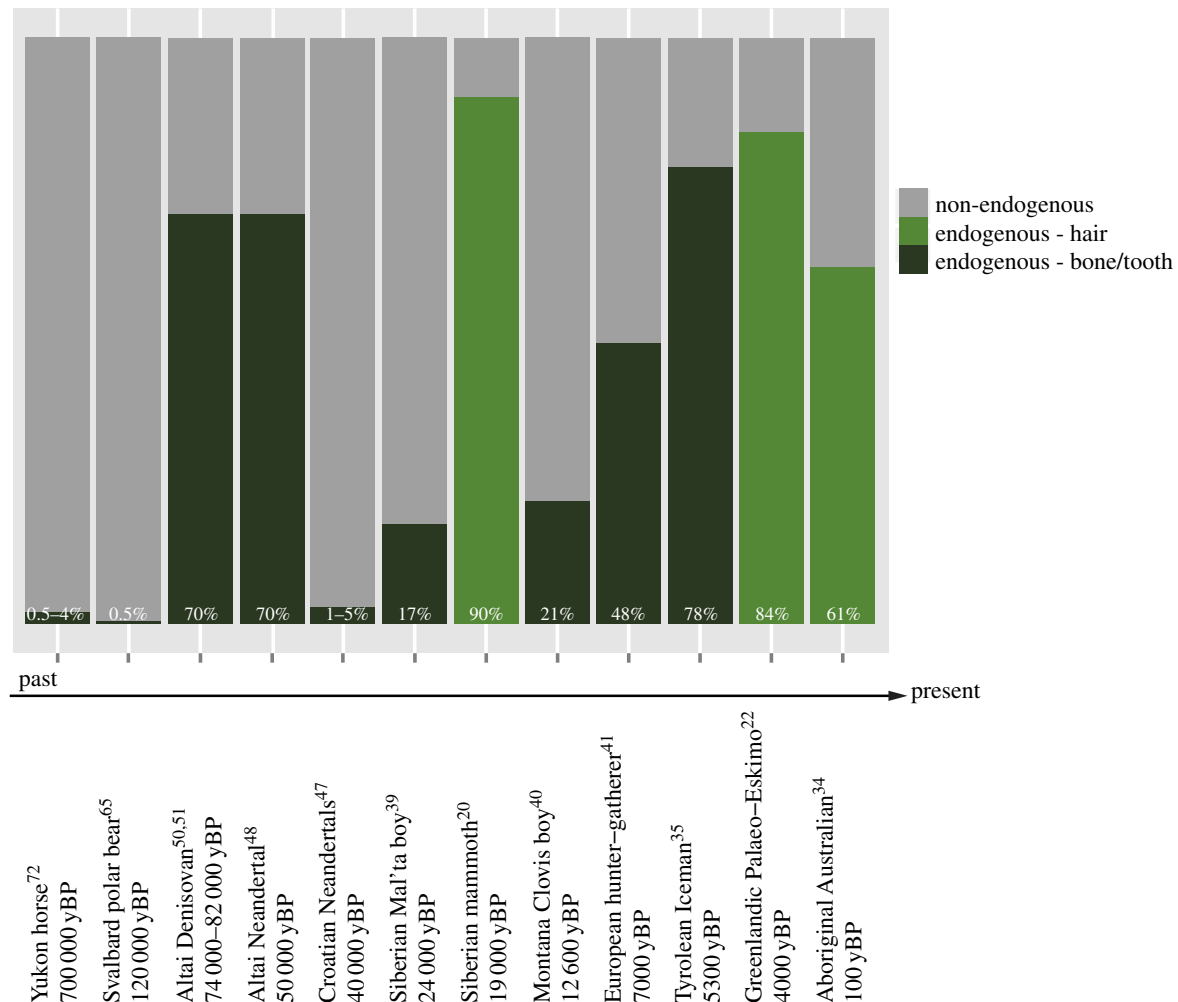


Figure 2. Endogenous content of ancient genomic extracts. Datasets are ordered from the most ancient sample to the most recent. 'yBP', years before present.

been engineered and shown to significantly increase amplification success from Pleistocene specimens, but their efficiency has not been investigated yet with next generation sequencing approaches [115].

(e) Increasing the relative amount of target DNA III: ancient DNA enrichment

The fact that most aDNA extracts show a minority of endogenous templates (figure 2) has led to the development of enrichment approaches aimed at reducing sequencing costs and improving the sequence quality of the targeted loci. Primer extension capture was the first of such methods and succeeded in recovering full mitochondrial sequence information from five Neandertal specimens [116] and from an approximately 30 kyr old modern human [117]. It was later superseded by other in-solution enrichment methods relying on biotinylated baits, either designed from known sequences and manufactured commercially [118], or prepared from modern DNA extracts [119]. These baits are subsequently used to target complementary library inserts. Both methods have shown great success in various aDNA contexts, often delivering complete mitochondrial genome sequences with high depth of coverage [42,72,117,119–123] and even pre-selected regions from the *Mycobacterium tuberculosis* genome [124]. The method has also been found to be relatively robust to the evolutionary distance separating probes and targets, yielding to significant enrichment despite

10–13% of sequence divergence [125]. Microarray-based hybridization capture has also performed well in enriching Neandertal DNA libraries [32] containing very low endogenous DNA content, and delivering the full bacterial genome of the causative agent of the mediaeval Black Death epidemic [106,126] as well as of historical leprosy strains [127]. One drawback of such approaches is that microarrays are designed from modern reference genomes. As a consequence, untargeted plasmids and/or loci potentially present in the historical strains and/or chromosomal rearrangements specific to the historical strain could remain undetected. This problem can be solved by using de novo genome assembly in cases where samples with exceptionally high pathogen DNA contents are available, such as for one of the three ancient leprosy samples that were genome sequenced [127].

Other types of enrichment approaches have been developed to target full human chromosomes [38], and even complete genomes [128,129], which performed well on poorly preserved DNA material. One such approach converts custom-designed microarray probes into an immortalized and amplifiable biotinylated library of baits that can be used for pulling-down orthologue inserts from aDNA libraries. This strategy enabled Fu *et al.* [38] to reconstruct all non-repetitive sequences of chromosome 21 in a 40 kyr old anatomically modern human from Tianyuan cave, China, from a set of immortalized baits recovered from nine microarrays and corresponding to 8.7 million probes across 30 Mb of the chromosome. A second approach achieves full genome enrichment in solution with

no prior need for microarray purchase, therefore cutting down on prohibitive microarray costs [128,129]. Here, modern DNA from a given organism is first built into a DNA library downstream of RNA polymerase T7-promoters so that genome-wide biotinylated RNA baits can be transcribed *in vitro*. Baits are then used to pull down orthologue inserts from regular aDNA libraries, followed by several washing steps to rid the library of exogenous DNA library inserts. This method has shown twofold to 13-fold enrichment in human DNA content, which provided enough SNP information for population assignment with minimal sequencing effort. Up to 19-fold enrichment was obtained on mammoth aDNA extracts, thus demonstrating that this approach can be applied on extinct species, even in absence of reference genomes [129]. Sequencing of genome-wide captured libraries is cost-effective, and thereby well suited for the analysis of large numbers of samples, which promises to move forward the field towards population genomics. aDNA whole genome enrichment and the development of capture methods targeting damaged DNA molecules will probably facilitate the characterization of Middle Pleistocene genomes in the near future.

9. What next?

Looking back 5 years, no one could have predicted the current state of ancient genomics. New sequencing technologies requiring no heavy infrastructure are being developed with the promise of delivering gigabases of sequence information at small cost. We can anticipate that ancient genomics will move on to the scale of population studies, with probable research areas in the reconstruction of human dispersal routes and demographic processes, such as the ones associated

with the Neolithic transition in Europe. In addition, we expect that palaeogenomics will soon help better understand the origins, evolution and pathogenicity of the bacterial and viral agents responsible for major historical pandemics in human history. Ancient genomics will also probably illuminate the domestication process by revealing the genes that have been artificially selected to transform wild animal and plant species into the variety of domesticated forms that we know today. Additionally, high-quality ancient genomes of megafaunal species together with genome-wide SNP surveys will document past demographic trajectories at unprecedented levels [130]. This type of approach will complete our current understanding of how species responded to major climatic changes in the past [80], a key to conservation genomics in the face of current global warming. Moreover, the accumulation of aDNA data will probably provide additional information on post-mortem DNA base modifications, which will be essential for understanding and correcting ancient genomic datasets, and also for reconstructing ancient methylation marks. Finally, we predict advances in the recovery of functional information from ancient specimens through proteomics [131], which has already delivered partial proteomes from extinct mammalian species [33,72], and through ancient epigenetic marks and nucleosome maps [28]. There is little doubt that, together, these prospects will catalyse yet another revolution in aDNA research.

Acknowledgements. We thank all members of the Centre for GeoGenetics for fruitful discussions.

Funding statement. This work was supported by the Danish National Research Foundation (DNRF94), two Marie-Curie Intra-European Fellowships IEF (299176 and 302617), the Marie-Curie Career Integration grant no. CIG-293845 and a Danish National Research Foundation FNU grant attributed to L.O.

References

- Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001 DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799. (doi:10.1093/nar/29.23.4793)
- Cooper A, Poinar HN. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139. (doi:10.1126/science.289.5482.1139b)
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. 2001 Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704–707. (doi:10.1038/35055536)
- Rogaev EI, Moliaka YK, Malyarchuk BA, Kondrashov FA, Derenko MV, Chumakov I, Grigorenko AP. 2006 Complete mitochondrial genome and phylogeny of Pleistocene mammoth. *PLoS Biol.* **4**, e73. (doi:10.1371/journal.pbio.0040073)
- Bon C *et al.* 2008 Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proc. Natl Acad. Sci. USA* **105**, 17 447–17 452. (doi:10.1073/pnas.0806143105)
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S. 1997 Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30. (doi:10.1016/S0092-8674(00)80310-4)
- Greenwood AD, Capelli C, Possnert G, Pääbo S. 1999 Nuclear DNA sequences from late Pleistocene megafauna. *Mol. Biol. Evol.* **16**, 1466–1473. (doi:10.1093/oxfordjournals.molbev.a026058)
- Poinar H, Kuch M, McDonald G, Martin P, Pääbo S. 2003 Nuclear gene sequences from a Late Pleistocene sloth coprolite. *Curr Biol.* **13**, 1150–1152. (doi:10.1016/S0960-9822(03)00450-0)
- Römpler H, Rohland N, Lalueza-Fox C, Willerslev E, Kuznetsova T, Rabeder G, Bertranpetit J, Schöneberg T, Hofreiter M. 2006 Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**, 62. (doi:10.1126/science.1128994)
- Krause J *et al.* 2006 Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* **439**, 724–727. (doi:10.1038/nature04432)
- Krause J *et al.* 2008 Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene–Pliocene boundary. *BMC Evol. Biol.* **8**, 220. (doi:10.1186/1471-2148-8-220)
- Tumpey TM *et al.* 2005 Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80. (doi:10.1126/science.1119392)
- Noonan JP *et al.* 2005 Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–599. (doi:10.1126/science.1113485)
- Noonan JP *et al.* 2006 Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118. (doi:10.1126/science.1131412)
- Margulies M *et al.* 2005 Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380. (doi:10.1038/nature03959)
- Poinar HN *et al.* 2006 Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394. (doi:10.1126/science.1123360)
- Gilbert MTP *et al.* 2007 Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927–1930. (doi:10.1126/science.1146971)
- Gilbert MT *et al.* 2008 Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl Acad. Sci. USA* **105**, 8327–8332. (doi:10.1073/pnas.0802315105)
- Gilbert MTP, Menez L, Janaway RC, Tobin DJ, Cooper A, Wilson AS. 2006 Resistance of degraded hair

- shafts to contaminant DNA. *Forensic Sci. Int.* **156**, 208–212. (doi:10.1016/j.forsciint.2005.02.021)
20. Miller W *et al.* 2008 Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387–390. (doi:10.1038/nature07446)
 21. Bentley DR *et al.* 2008 Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59. (doi:10.1038/nature07517)
 22. Rasmussen M *et al.* 2010 Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762. (doi:10.1038/nature08835)
 23. Alexander DH, Novembre J, Lange K. 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664. (doi:10.1101/gr.094052.109)
 24. Patterson N, Price AL, Reich D. 2006 Population structure and eigenanalysis. *PLoS Genet.* **2**, e190. (doi:10.1371/journal.pgen.0020190)
 25. Willerslev E, Cooper A. 2005 Ancient DNA. *Proc. R. Soc. B* **272**, 3–16. (doi:10.1098/rspb.2004.2813)
 26. Green RE *et al.* 2006 Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336. (doi:10.1038/nature05336)
 27. Wall JD, Kim SK. 2007 Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genetics* **3**, e175. (doi:10.1371/journal.pgen.0030175)
 28. Pedersen JS *et al.* 2013 Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **24**, 454–466. (doi:10.1101/gr.163592.113)
 29. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. 2010 Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87. (doi:10.1093/nar/gkp1163)
 30. Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, Suter CM. 2012 High-resolution analysis of cytosine methylation in ancient DNA. *PLoS ONE* **7**, e30226. (doi:10.1371/journal.pone.0030226)
 31. Fogg MJ, Pearl LH, Connolly BA. 2002 Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol.* **9**, 922–927. (doi:10.1038/nsb867)
 32. Burbano HA *et al.* 2010 Targeted investigation of the Neanderthal genome by array-based sequence capture. *Science* **328**, 723–725. (doi:10.1126/science.1188046)
 33. Cappellini E *et al.* 2012 Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* **11**, 917–926. (doi:10.1021/pr200721u)
 34. Rasmussen M *et al.* 2011 An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98. (doi:10.1126/science.1211177)
 35. Keller A *et al.* 2012 New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698. (doi:10.1038/ncomms1701)
 36. Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MT, Götherström A, Jakobsson M. 2012 Origins and genetic legacy of Neolithic farmers and hunter–gatherers in Europe. *Science* **336**, 466–469. (doi:10.1126/science.1216304)
 37. Sánchez-Quinto F *et al.* 2012 Genomic affinities of two 7,000-year-old Iberian hunter–gatherers. *Curr. Biol.* **22**, 1494–1499. (doi:10.1016/j.cub.2012.06.005)
 38. Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, Pääbo S. 2013 DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227. (doi:10.1073/pnas.1221359110)
 39. Raghavan M *et al.* 2013 Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91. (doi:10.1038/nature12736)
 40. Rasmussen M *et al.* 2014 The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229. (doi:10.1038/nature13025)
 41. Olalde I *et al.* 2014 Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228. (doi:10.1038/nature12960)
 42. Fu Q *et al.* 2013 A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559. (doi:10.1016/j.cub.2013.02.044)
 43. Cavalli-Sforza LL, Menozzi R, Piazza A. 1994 *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
 44. HUGO Pan-Asian SNP Consortium. 2009 Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545. (doi:10.1126/science.1177074)
 45. Pickrell JK, Pritchard JK. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967. (doi:10.1371/journal.pgen.1002967)
 46. Schurr TG, Sherry ST. 2004 Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am. J. Hum. Biol.* **16**, 420–439. (doi:10.1002/ajhb.20041)
 47. Green RE *et al.* 2010 A draft sequence of the Neanderthal genome. *Science* **328**, 710–722. (doi:10.1126/science.1188021)
 48. Prüfer K *et al.* 2013 The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49. (doi:10.1038/nature12886)
 49. Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP, Pääbo S. 2010 The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897. (doi:10.1038/nature08976)
 50. Reich D *et al.* 2010 Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature* **468**, 1053–1060. (doi:10.1038/nature09710)
 51. Meyer M *et al.* 2012 A high-coverage genome sequence from an Archaic Denisovan individual. *Science* **338**, 222–226. (doi:10.1126/science.1224344)
 52. Meyer M *et al.* 2013 A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406. (doi:10.1038/nature12788)
 53. Orlando L. 2014 A 400,000-year-old mitochondrial genome questions phylogenetic relationships amongst archaic hominins. *BioEssays* **36**, 598–605. (doi:10.1002/bies.201400018)
 54. Durand EY, Patterson N, Reich D, Slatkin M. 2011 Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252. (doi:10.1093/molbev/msr048)
 55. Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. 2012 The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947. (doi:10.1371/journal.pgen.1002947)
 56. Yang MA, Malaspina AS, Durand EY, Slatkin M. 2012 Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Mol. Biol. Evol.* **29**, 2987–2995. (doi:10.1093/molbev/mss117)
 57. Eriksson A, Manica A. 2012 Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl Acad. Sci. USA* **109**, 13 956–13 960. (doi:10.1073/pnas.1200567109)
 58. Eriksson A, Manica A. 2014 The doubly conditioned frequency spectrum does not distinguish between ancient population structure and hybridization. *Mol. Biol. Evol.* **31**, 618–621. (doi:10.1093/molbev/msu103)
 59. Reich D *et al.* 2011 Denisova admixture and the first modern human dispersals into southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528. (doi:10.1016/j.ajhg.2011.09.005)
 60. Skoglund P, Jakobsson M. 2011 Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18 301–18 306. (doi:10.1073/pnas.1108181108)
 61. Lopez-Valenzuela M, Ramirez O, Rosas A, Garcia-Vargas S, de la Rasilla M, Lalueza-Fox C, Espinosa-Parrilla Y. 2012 An ancestral miR-1304 allele present in Neanderthals regulates genes involved in enamel formation and could explain dental differences with modern humans. *Mol. Biol. Evol.* **29**, 1797–1806. (doi:10.1093/molbev/mss023)
 62. Mendez FL, Watkins JC, Hammer MF. 2012 A haplotype at STAT2 introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91**, 265–274. (doi:10.1016/j.ajhg.2012.06.015)
 63. Mendez FL, Watkins JC, Hammer MF. 2012 Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. *Mol. Biol. Evol.* **29**, 1513–1520. (doi:10.1093/molbev/msr301)
 64. Abi-Rached L *et al.* 2011 The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94. (doi:10.1126/science.1209202)
 65. Miller W *et al.* 2012 Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl Acad. Sci. USA* **109**, 12 000–12 005. (doi:10.1073/pnas.1200000109)

- Sci. USA* **109**, E2382–E2390. (doi:10.1073/pnas.1210506109)
66. Valdiosera C, García N, Dalén L, Smith C, Kahle RD, Lidén K, Angerbjörn A, Arsuaga JL, Götherström A. 2006 Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. *Biol. Lett.* **22**, 601–603. (doi:10.1098/rsbl.2006.0515)
 67. Willerslev E *et al.* 2007 Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**, 111–114. (doi:10.1126/science.1141758)
 68. Willerslev E, Hansen AJ, Rønn R, Brand TB, Barnes I, Wiuf C, Gilchinsky D, Mitchell D, Cooper A. 2004 Long-term persistence of bacterial DNA. *Curr. Biol.* **14**, R9–R10. (doi:10.1016/j.cub.2003.12.012)
 69. Stewart JR, Stringer CB. 2012 Human evolution out of Africa: the role of refugia and climate change. *Science* **335**, 1317–1321. (doi:10.1126/science.1215627)
 70. Barnosky AD, Bell CJ, Emslie SD, Goodwin HT, Mead JL, Repenning CA, Scott E, Shabel AB. 2004 Exceptional record of mid-Pleistocene vertebrates helps differentiate climatic from anthropogenic ecosystem perturbations. *Proc. Natl Acad. Sci. USA* **101**, 9297–9302. (doi:10.1073/pnas.0402592101)
 71. Sardella R. 2007 Vertebrate records, Mid-Pleistocene of Europe. In *Encyclopedia of quaternary science*, vol. 4 (ed. SA Elias), pp. 3224–3232. Amsterdam, The Netherlands: Elsevier.
 72. Orlando L *et al.* 2013 Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78. (doi:10.1038/nature12323)
 73. Storer JE. 2004 A middle Pleistocene (late Irvingtonian) mammalian fauna from Thistle Creek, Klondike Goldfields region of Yukon Territory, Canada. *Paludicola* **4**, 137–150.
 74. Westgate JA *et al.* 2009 Gold Run tephra: a Middle Pleistocene stratigraphic and paleoenvironmental marker across west-central Yukon Territory, Canada. *Can. J. Earth Sci.* **46**, 465–478. (doi:10.1139/E09-029)
 75. Preece SJ, Westgate JA, Froese DG, Pearce NJG, Perkins WT. 2011 A catalogue of late Cenozoic tephra beds in the Klondike goldfields and adjacent areas, Yukon Territory. Yukon Geological Survey Contribution. *Can. J. Earth Sci.* **48**, 1386–1418. (doi:10.1139/e10-110)
 76. Allentoft ME *et al.* 2012 The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B* **279**, 4724–4733. (doi:10.1098/rspb.2012.1745)
 77. Orlando L *et al.* 2011 True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* **21**, 1705–1719. (doi:10.1101/gr.122747.111)
 78. Ginolhac A *et al.* 2012 Improving the performance of true single molecule sequencing for ancient DNA. *BMC Genomics* **13**, 177. (doi:10.1186/1471-2164-13-177)
 79. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, Krogh A, Orlando L. 2012 Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178. (doi:10.1186/1471-2164-13-178)
 80. Lorenzen ED *et al.* 2011 Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364. (doi:10.1038/nature10574)
 81. Millar CD, Lambert DM. 2013 Ancient DNA: towards a million-year-old genome. *Nature* **499**, 34–35. (doi:10.1038/nature12263)
 82. Dabney J *et al.* 2014 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15 758–15 763. (doi:10.1073/pnas.1314445110)
 83. Gansauge MT, Meyer M. 2013 Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748. (doi:10.1038/nprot.2013.038)
 84. Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, Doebley J, Pääbo S. 2003 Early allelic selection in maize as revealed by ancient DNA. *Science* **302**, 1206–1208. (doi:10.1126/science.1089056)
 85. Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. 2009 Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* **4**, e6301. (doi:10.1371/journal.pone.0006301)
 86. Palmer SA, Clapham AJ, Rose P, Freitas FO, Owen BD, Beresford-Jones D, Moore JD, Kitchen JL, Allaby RG. 2012 Archaeogenomic evidence of punctuated genome evolution in *Gossypium*. *Mol. Biol. Evol.* **29**, 2031–2038. (doi:10.1093/molbev/mss070)
 87. Li C, Lister DL, Li H, Xu Y, Cui Y, Bower MA, Jones MK, Zhou H. 2011 Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *J. Arc. Sci.* **38**, 115–119. (doi:10.1016/j.jas.2010.08.016)
 88. Kistler L, Montenegro A, Smith BD, Gifford JA, Green RE, Newsom LA, Shapiro B. 2014 Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proc. Natl Acad. Sci. USA* **111**, 2937–2941. (doi:10.1073/pnas.1318678111)
 89. Wales N, Allaby R, Gilbert MTP, Willerslev E. 2013 Ancient plant DNA. In *Encyclopedia of quaternary science*, vol. 2 (ed. SA Elias), pp. 705–715, 2nd edn. Amsterdam, The Netherlands: Elsevier.
 90. Martin MD *et al.* 2013 Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* **4**, 2172. (doi:10.1038/ncomms3172)
 91. Yoshida K *et al.* 2013 The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* **2**, e00731. (doi:10.7554/eLife.01108)
 92. Martin MD, Ho SY, Wales N, Ristaino J, Gilbert MTP. 2014 Persistence of the mitochondrial lineage responsible for the Irish potato famine in extant New World *Phytophthora infestans*. *Mol. Biol. Evol.* **31**, 1414–1420. (doi:10.1093/molbev/msu086)
 93. Parducci L *et al.* 2012 Glacial survival of boreal trees in northern Scandinavia. *Science* **335**, 1083–1086. (doi:10.1126/science.1216043)
 94. Fordyce SL *et al.* 2013 Deep sequencing of RNA from ancient maize kernels. *PLoS ONE* **8**, e50961. (doi:10.1371/journal.pone.0050961)
 95. Oskam CL *et al.* 2010 Fossil avian eggshell preserves ancient DNA. *Proc. R. Soc. B* **277**, 1991–2000. (doi:10.1098/rspb.2009.2019)
 96. Adler CJ *et al.* 2013 Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455. (doi:10.1038/ng.2536)
 97. Warinner C *et al.* 2014 Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344. (doi:10.1038/ng.2906)
 98. Briggs AW *et al.* 2007 Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14 616–14 621. (doi:10.1073/pnas.0704665104)
 99. Seguin-Orlando A, Schubert M, Clary J, Stagegaard J, Alberdi MT, Prado JL, Prieto A, Willerslev E, Orlando L. 2013 Ligation bias in Illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS ONE* **8**, e78575. (doi:10.1371/journal.pone.0078575)
 100. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012 Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131. (doi:10.1371/journal.pone.0034131)
 101. Overballe-Petersen S, Orlando L, Willerslev E. 2012 Next-generation sequencing offers new insights into DNA degradation. *Trends Biotech.* **30**, 364–368. (doi:10.1016/j.tibtech.2012.03.007)
 102. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, Jakobsson M. 2014 Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234. (doi:10.1073/pnas.1318934111)
 103. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011 *MAPDAMAGE*: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153–2155. (doi:10.1093/bioinformatics/btr347)
 104. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013 *MAPDAMAGE20*: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684. (doi:10.1093/bioinformatics/btt193)
 105. Green RE *et al.* 2008 A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426. (doi:10.1016/j.cell.2008.06.021)
 106. Schuenemann VJ *et al.* 2011 Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752. (doi:10.1073/pnas.1105107108)
 107. Bos KI *et al.* 2011 A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510. (doi:10.1038/nature10549)

108. Rohland N, Hofreiter M. 2007 Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762. (doi:10.1038/nprot.2007.247)
109. Deagle BE, Eveson JP, Jarman SN. 2006 Quantification of damage in DNA recovered from highly degraded samples: a case study on DNA in faeces. *Front. Zool.* **3**, 11. (doi:10.1186/1742-9994-3-11)
110. Salamon M, Tuross N, Arensburg B, Weiner S. 2005 Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc. Natl Acad. Sci. USA* **102**, 13 783–13 788. (doi:10.1073/pnas.0503718102)
111. Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J, Poinar H. 2009 New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* **37**, 3215–3229. (doi:10.1093/nar/gkp159)
112. Der Sarkissian C, Ermini L, Jónsson H, Alekseev AN, Crubezy E, Shapiro B, Orlando L. 2014 Shotgun microbial profiling of fossil remains. *Mol. Ecol.* **23**, 1780–1798. (doi:10.1111/mec.12690)
113. Dabney J, Meyer M. 2012 Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94. (doi:10.2144/000113809)
114. Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, Meyer M. 2010 Road blocks on paleogenomes—polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Res.* **38**, e161. (doi:10.1093/nar/gkq572)
115. d'Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, Cadet J, Woodgate R, Pääbo S, Holliker P. 2007 Molecular breeding of polymerases for amplification of ancient DNA. *Nat. Biotechnol.* **25**, 939–943. (doi:10.1038/nbt1321)
116. Briggs AW *et al.* 2009 Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321. (doi:10.1126/science.1174462)
117. Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Dereviako A, Pääbo S. 2010 A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236. (doi:10.1016/j.cub.2009.11.068)
118. Avila-Arcos MC *et al.* 2011 Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **1**, 74. (doi:10.1038/srep00074)
119. Maricic T, Whitten M, Pääbo S. 2010 Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004. (doi:10.1371/journal.pone.0014004)
120. Vilstrup JT *et al.* 2013 Mitochondrial phylogenomics of modern and ancient equids. *PLoS ONE* **8**, e55950. (doi:10.1371/journal.pone.0055950)
121. Horn S. 2011 Case study: enrichment of ancient mitochondrial DNA by hybridization capture. In *Ancient DNA: methods and protocols* (eds B Shapiro, M Hofreiter), pp. 189–195. Totowa, NJ: Humana Press.
122. Zhang H *et al.* 2013 Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nat. Commun.* **4**, 2755. (doi:10.1038/ncomms3755)
123. Thalmann O *et al.* 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874. (doi:10.1126/science.1243650)
124. Bouwman AS, Kennedy SL, Muller R, Stephens RH, Holst M, Caffell AC, Roberts CA, Brown TA. 2012 Genotype of a historic strain of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **109**, 18 511–18 516. (doi:10.1073/pnas.1209444109)
125. Mason VC, Li G, Helgen KM, Murphy WJ. 2011 Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Res.* **21**, 1695–1704. (doi:10.1101/gr.120196.111)
126. Bos KI, Stevens P, Nieselt K, Poinar HN, Dewitte SN, Krause J. 2012 *Yersinia pestis*: new evidence for an old infection. *PLoS ONE* **7**, e49803. (doi:10.1371/journal.pone.0049803)
127. Schuenemann VJ *et al.* 2013 Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183. (doi:10.1126/science.1238286)
128. Carpenter ML *et al.* 2013 Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* **93**, 852–864. (doi:10.1016/j.ajhg.2013.10.002)
129. Enk JM, Devault AM, Kuch M, Murgu YE, Rouillard JM, Poinar HN. 2014 Ancient whole genome enrichment using baits built from modern DNA. *Mol. Biol. Evol.* **31**, 1292–1294. (doi:10.1093/molbev/msu074)
130. Mourier T, Ho SY, Gilbert MT, Willerslev E, Orlando L. 2012 Statistical guidelines for detecting past population shifts using ancient DNA. *Mol. Biol. Evol.* **29**, 2241–2251. (doi:10.1093/molbev/mss094)
131. Cappellini E, Collins MJ, Gilbert MT. 2014 Biochemistry unlocking ancient protein palimpsests. *Science* **343**, 1320–1322. (doi:10.1126/science.1249274)